SDSC HPC Resources

DataStar: A 15.6 TF Power4 + Federation Switch System
Blue Gene: 17.1 TF 6144 PowerPC processors

Amit Majumdar

(majumdar@sdsc.edu)

and

Mahidhar Tatineni

(mahidhar@sdsc.edu)

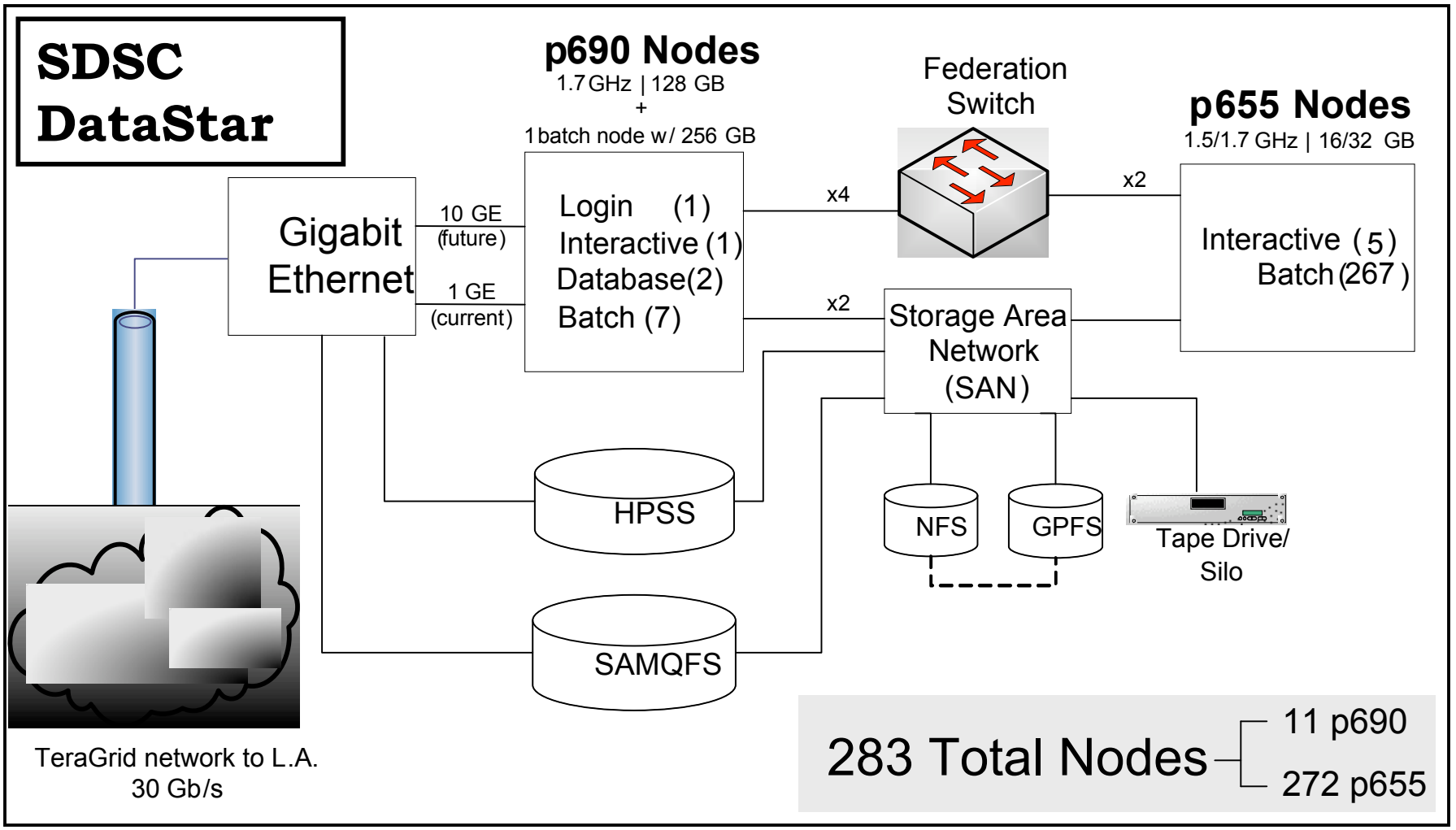
ASC Center Visits, October 2006

SDSC Resources

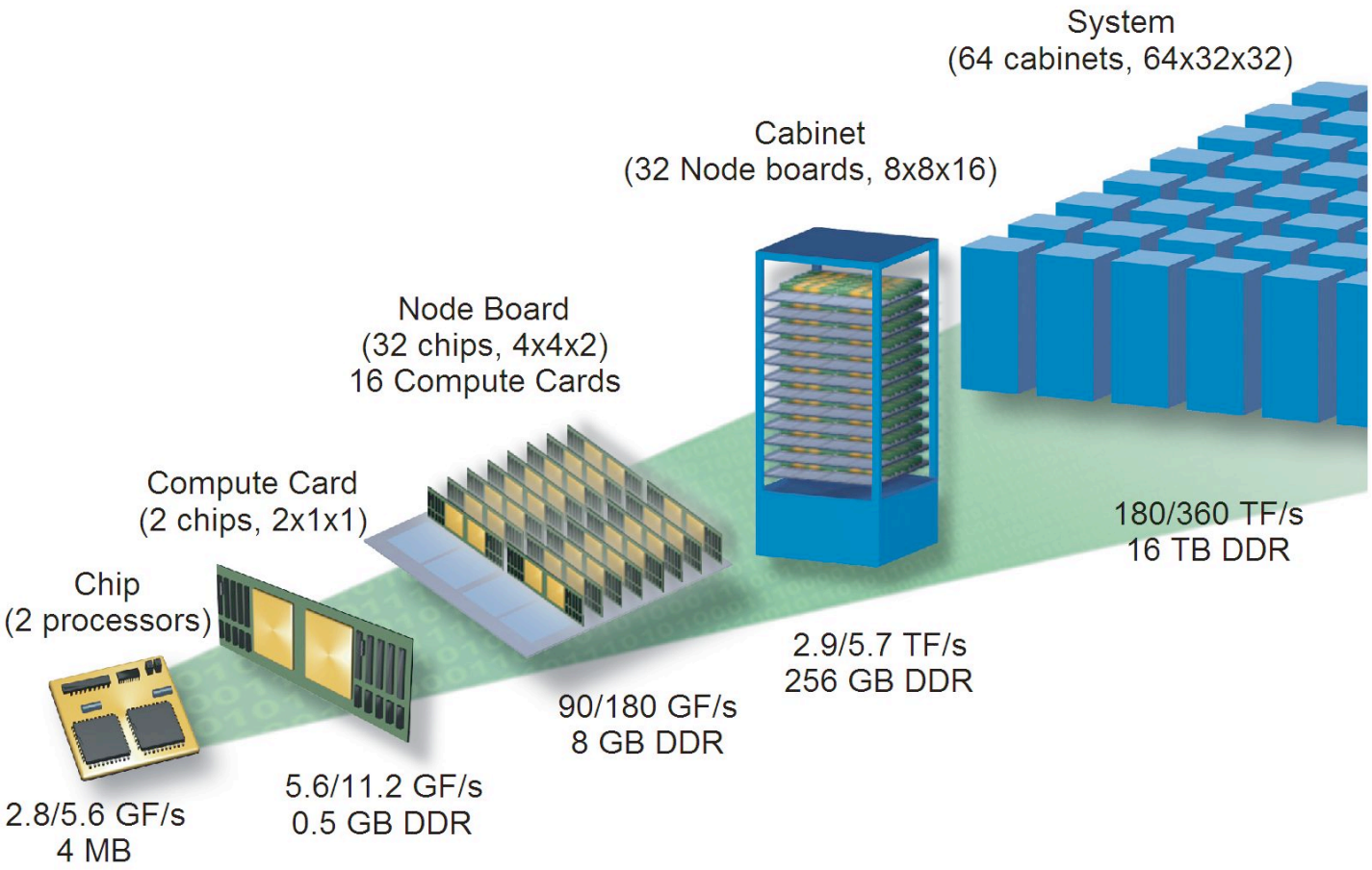
- **A NSF center with compute and data resources allocated *freely* through peer review process**
- **One of the emphasis of SDSC for national Cyberinfrastructure initiative is Data intensive computing**
- **SDSC DataStar with 2528 power 4+ processors (15.6 TF), federation switch, 115 TB GPFS.**
- **SDSC Blue Gene will be upgraded to three racks with 6144 processors (17.1 TF total). ASC users can transfer their allocation from DataStar to BlueGene. One SU on DataStar will be transferred as 2.5 SUs on the Blue Gene.**
- **225 TB GPFS-WAN filesystem shared between all SDSC production systems.**

DataStar Configuration

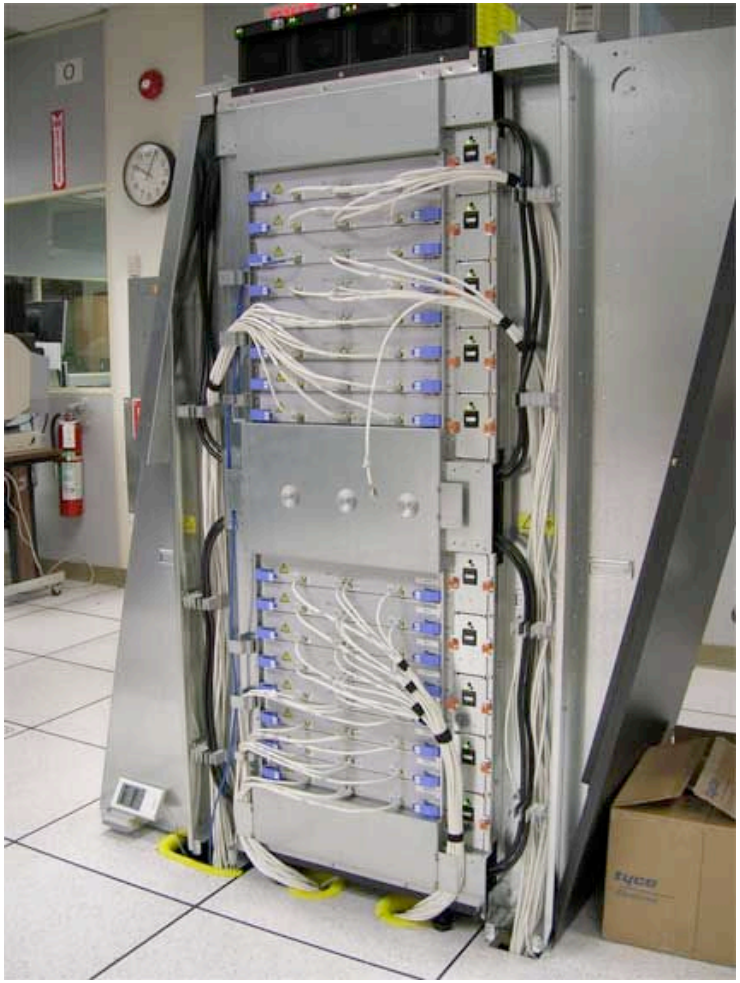
- **15.6 TF, 2528 processors total**
- **11 32-way 1.7 GHz IBM p690**
  - 2 nodes 64 GB memory for login and system use
  - 4 nodes 128 GB memory + 1 node 256 GB for batch scientific computation
  - 3 nodes 128 GB memory for database, DiscoveryLink, HPSS
  - 1 node 256 GB memory for interactive use (post processing, visualization)
- **176 8-way 1.5 GHz IBM p655**
  - 16 GB memory
  - Batch scientific computation
- **96 8-way 1.7 GHz IBM p655**
  - 32 GB memory
  - Batch scientific computation
- **All nodes Federation switch attached**
- **All nodes SAN attached**
- **Parallel filesystem: 115 TB GPFS; 225 TB GPFS-WAN (Shared with Blue Gene, TG IA-64 cluster)**


SDSC DataStar
p690 Nodes
1.7 GHz | 128 GB
+
1 batch node w/ 256 GB
Federation Switch
p655 Nodes
1.5/1.7 GHz | 16/32 GB
Gigabit Ethernet
10 GE (future)
1 GE (current)
Login (1)
Interactive (1)
Database(2)
Batch (7)
x4
x2
Interactive (5)
Batch (267)
x2
Storage Area Network (SAN)
HPSS
NFS
GPFS
Tape Drive/ Silo
SAMQFS
TeraGrid network to L.A.
30 Gb/s
283 Total Nodes
11 p690
272 p655

Blue Gene System Overview: Chips to Racks



Blue Gene System Overview: SDSC's three-rack system



Soon to be three racks!

BG System Overview: SDSC's three-rack system

- **3072 compute nodes & 384 I/O nodes (each with 2p)**
  - Most I/O-rich configuration possible (8:1 compute:I/O node ratio)
  - Identical hardware in each node type with different networks wired
  - Compute nodes connected to: torus, tree, global interrupt, & JTAG
  - I/O nodes connected to: tree, global interrupt, Gigabit Ethernet, & JTAG
- **Two half racks (also confusingly called midplanes)**
  - Connected via link chips
- **Front-end nodes (4 B80s (4 procs each) + 1 Power 5 node (4procs))**
- **Service node (p275 with 2p)**

BG System Overview: Processor Chip (= System-on-a-chip)

- **Two 700-MHz PowerPC 440 processors**
  - Each with two floating-point units
  - Each with 32-kB L1 data caches that are noncoherent
  - 4 flops/proc-clock peak (=2.8 Gflops/proc)
  - 2 8-B loads or stores / proc-clock peak in L1 (=11.2 GBps/proc)
- **Shared 2-kB L2 cache (or prefetch buffer)**
- **Shared 4-MB L3 cache**
- **Five network controllers (though not all wired to each node)**
  - 3-D torus (for point-to-point MPI operations: 175 MBps nom x 6 links x 2 ways)
  - Tree (for most collective MPI operations: 350 MBps nom x 3 links x 2 ways)
  - Global interrupt (for MPI_Barrier: low latency)
  - Gigabit Ethernet (for I/O)
  - JTAG (for machine control)
- **Memory controller for 512 MB of off-chip, shared memory**

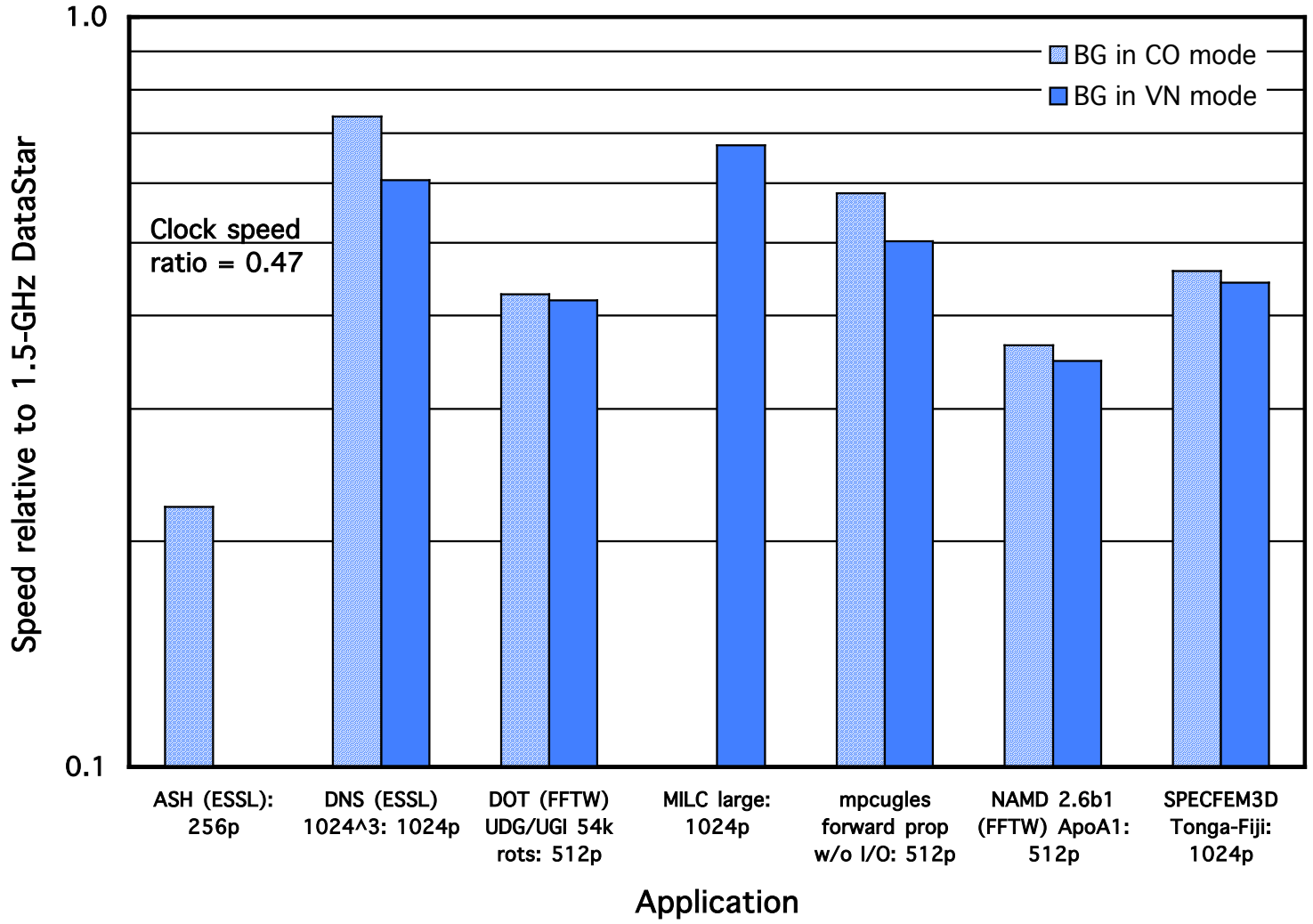
BG System Overview: Multiple operating systems & functions

- **Compute nodes: run Compute Node Kernel (CNK = blrts)**
  - Each run only one job at a time
  - Each use very little memory for CNK
- **I/O nodes: run Embedded Linux**
  - Run CIOD to manage compute nodes
  - Perform file I/O
  - Run GPFS
- **Front-end nodes: run SuSE SLES9 Linux/PPC64**
  - Support user logins
  - Run cross compilers & linker
  - Run parts of mpirun to submit jobs & LoadLeveler to manage jobs
- **Service node: runs SuSE SLES9 Linux/PPC64**
  - Uses DB2 to manage four system databases
  - Runs control system software, including MMCS
  - Runs other parts of mpirun & LoadLeveler
- **Currently running V1R3 version of driver. All libraries need to be recompiled for the new driver. This has already been completed for the libraries in /usr/local/apps.**

Major applications ported and being run on BG at SDSC span various disciplines

Code name	Discipline	Description	Implementors
ASH (ESSL)	Astrophysics	3-D solar convection	Ben Brown (Colorado)
DNS (ESSL)	Engineering	Direct numerical simulatio	Diego Donzis (Georgia Tech) &
		of 3-D turbulence	Dmitry Pekurovsky (SDSC)
DOT (FFTW)	Biophysics	Protein docking	Susan Lindsey (SDSC) &
			Wayne Pfeiffer (SDSC)
MILC *	Physics	Quantum chromodynamic	Doug Toussaint (Arizona)
mpcugles	Engineering	3-D fluid dynamics	Krishnan Mahesh (UMinn)
			Giri Chukkapalli (SDSC)
NAMD 2.6b1 *	Biophysics	Molecular dynamics	Sameer Kumar (IBM)
(FFTW)			
SPECFEM3D	Geophysics	3-D seismic wave	Brian Savage
		propagation	(Carnegie Institution)
SDSC Blue Gene went into production Oct '05			

Speed of BG relative to DataStar is close to clock speed ratio (=0.7/1.5) for several applications on 512p or 1024p; CO & VN mode perform similarly (per MPI p)



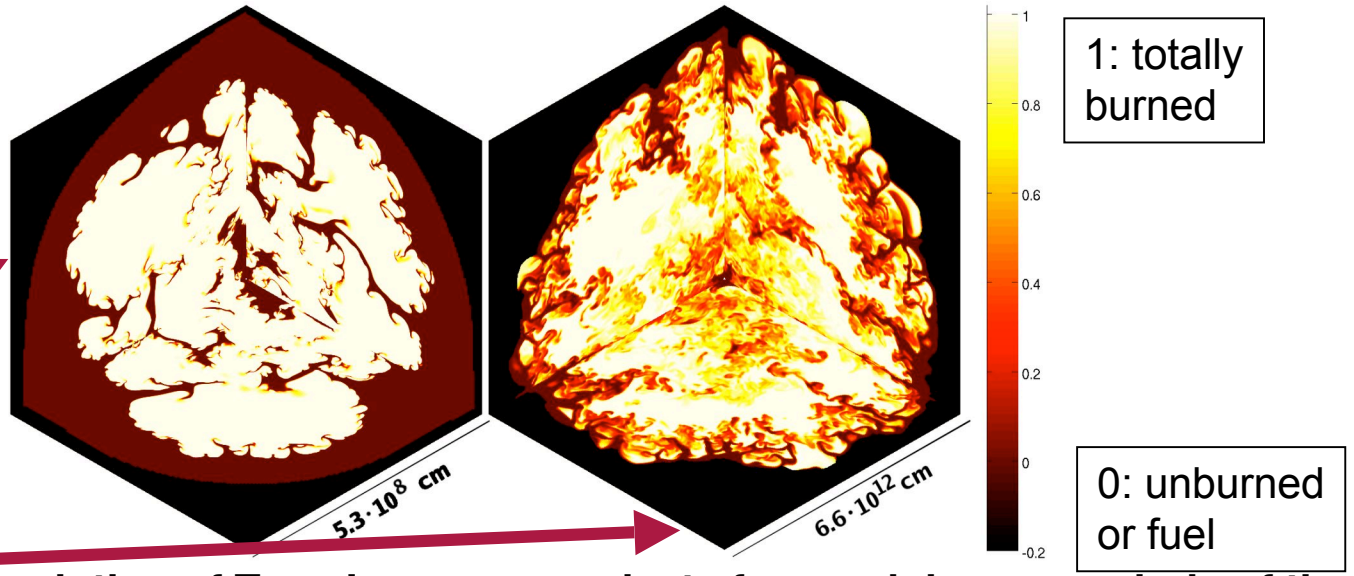
Longest-ever Simulation of Type Ia Supernova

Alexei Poludnenko, Alexei Khokhlov, Don Lamb – U. Chicago

- **The first self-consistent 3-D numerical simulation of the Type Ia supernova deflagration explosion from the moment of ignition through the active explosion phase and followed up to the period of 11 days**
- **The current state of the art multidimensional models of such astrophysical phenomena have typically followed the evolution of the system for a few tens of seconds**

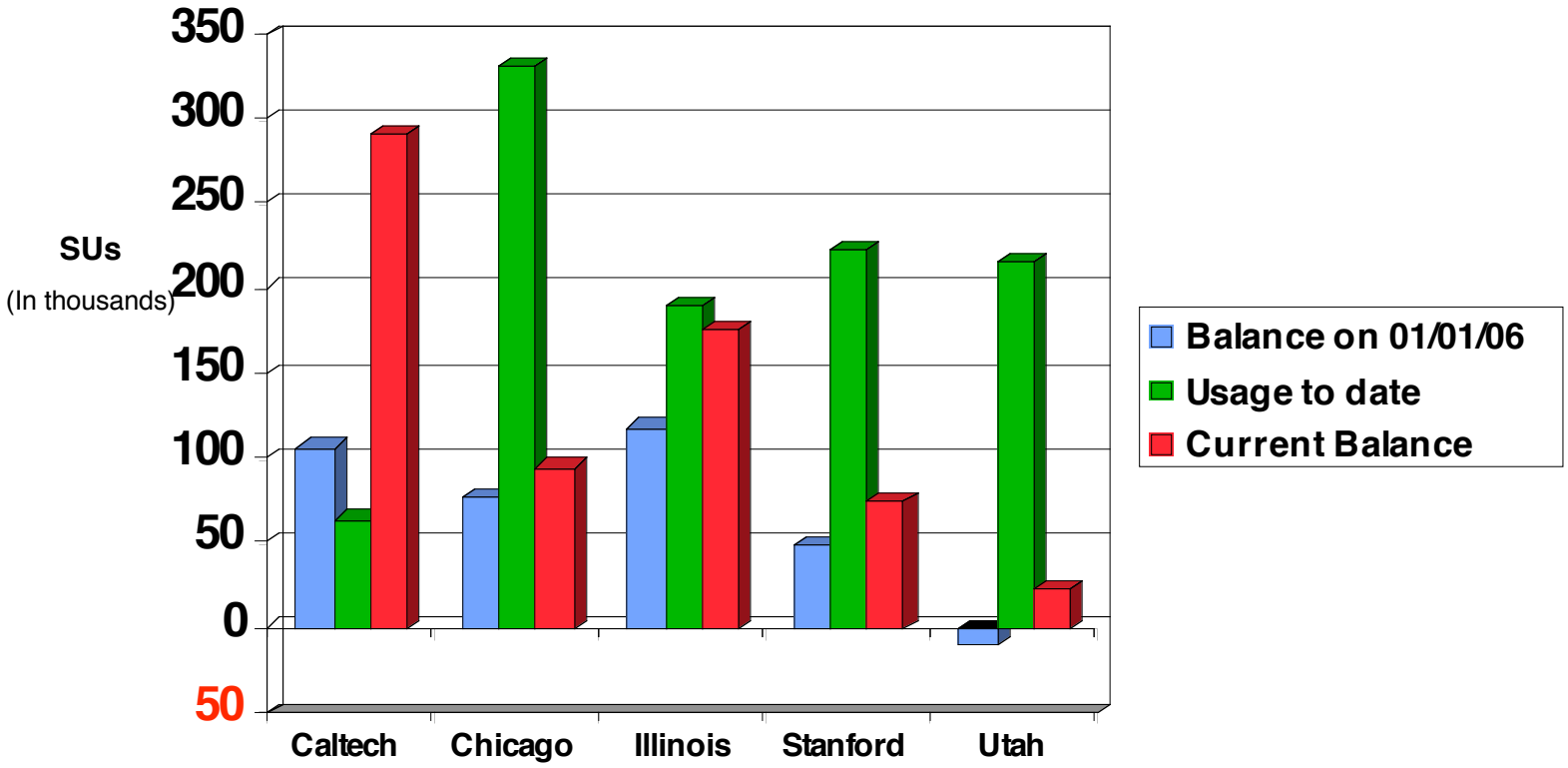
Flame structure in the star at 2 sec

and at 77 min



- **Post-explosion evolution of Type Ia supernova lasts for much longer periods of time going through various stages with different physical processes being important at different stages**
- **On 512 DS processors - total SU usage in August was ~30,000; Overall SU, included development & testing of the numerical code, was ~200,000 SUs**

ASC: Current Usage at SDSC



SDSC Resources and Links

- **SDSC's User Services page provides online user guides, frequently updated user news, consultant support and training:**
  http://www.sdsc.edu/user_services/

- **The DataStar userguide is located online at:**
  http://www.sdsc.edu/user_services/datastar/

- **The Blue Gene userguide is located online at:**
  http://www.sdsc.edu/user_services/bluegene/

- **For SDSC consulting information please visit:**
  http://www.sdsc.edu/user_services/consulting/

- **Submit a Ticket or E-mail consult@sdsc.edu (M - F 5:00am - 5:00pm PST).**

- **For technical support issues that cannot be submitted electronically, call 1-866-336-2357 (M - F 9:00am - 5:00pm PST).**